

## Profile Likelihood Confidence Intervals for GLM's

The standard procedure for computing a confidence interval (CI) for a parameter in a generalized linear model is by the formula: estimate  $\pm$  percentile  $\times$  SE(estimate), where SE is the standard error. The percentile is selected according to a desired confidence level and a reference distribution (a t-distribution for regression coefficients in a linear model, otherwise a standard normal distribution). This procedure is commonly referred to as a Wald-type CI. It may work poorly if the distribution of the parameter estimator is markedly skewed or if the standard error is a poor estimate of the standard deviation of the estimator. Since the standard errors in GLM's are based on asymptotic variances obtained from the information matrix, Wald CI's may perform poorly for small to moderate sample sizes. Profile likelihood confidence intervals don't assume normality of the estimator and appear to perform better for small samples sizes than Wald CI's. They are, nonetheless, still based on an asymptotic approximation – the asymptotic chi-square distribution of the log likelihood ratio test statistic.

### Profile likelihood method

Consider a model with parameters  $\theta$  and  $\delta$  where  $\theta$  is the parameter of interest and  $\delta$  is the (vector of) additional parameter(s) in the model. For example,  $\theta$  could be one of the  $\beta_j$ 's in a GLM and  $\delta$  is the remaining  $\beta_j$ 's. Denote by  $L(\theta, \delta)$  the likelihood function. The profile likelihood function for  $\theta$  is

$$L_1(\theta) = \max_{\delta} L(\theta, \delta).$$

For each value of  $\theta$ ,  $L_1(\theta)$  is the maximum of the likelihood function over the remaining parameters. Thus, the profile likelihood function is not a likelihood function; each point on the profile likelihood function is the maximum value of a likelihood function.

The idea of a profile likelihood confidence interval is to invert a likelihood-ratio test to obtain a CI for the parameter in question. A  $100(1 - \alpha)\%$  confidence interval for  $\theta$  is the set of all values  $\theta_0$  such that a two-sided test of the null hypothesis  $H_0 : \theta = \theta_0$  would not be rejected at the  $\alpha$  level of significance. The likelihood ratio test statistic of the hypothesis  $H_0 : \theta = \theta_0$  (where  $\theta_0$  is a fixed value) equals the difference between  $2 \log L$  for the “full” model and  $2 \log L$  for the reduced model which has  $\theta$  fixed at  $\theta_0$ , *i.e.*,  $2[\log L(\hat{\theta}, \hat{\delta}) - \log L(\theta_0, \hat{\delta}_0)] = 2[\log L(\hat{\theta}, \hat{\delta}) - \log L_1(\theta_0)]$ , where  $\hat{\theta}$  and  $\hat{\delta}$  are the MLE's for the full model and  $\hat{\delta}_0$  is the MLE of  $\delta$  for the reduced model with  $\theta = \theta_0$ . Based on the asymptotic chi-square distribution of the likelihood ratio test statistic if the null hypothesis is true, the test of  $H_0 : \theta = \theta_0$  will not be rejected at the  $\alpha$  level of significance if and only if

$$2[\log L(\hat{\theta}, \hat{\delta}) - \log L(\theta_0, \hat{\delta}_0)] = 2[\log L(\hat{\theta}, \hat{\delta}) - \log L_1(\theta_0)] < \chi_{1-\alpha}^2(1),$$

or, equivalently, if and only if

$$\log L_1(\theta_0) > \log L(\hat{\theta}, \hat{\delta}) - \chi_{1-\alpha}^2(1)/2,$$

where  $\chi_{1-\alpha}^2(1)$  is the  $1 - \alpha$  quantile of a  $\chi^2$  distribution with 1 d.f. Since  $\log L(\hat{\theta}, \hat{\delta})$  is fixed, we can plot the profile log-likelihood function  $\log L_1(\theta)$  and simply look at the interval for which it exceeds  $\log L(\hat{\theta}, \hat{\delta}) - \chi_{1-\alpha}^2(1)/2$ . That's the  $100(1 - \alpha)\%$  confidence interval for  $\theta$ .

The R function `confint(m)` where `m` is a fitted GLM computes likelihood profile confidence intervals for all the parameters in the model. The example below illustrates the use of this command as well as how to compute and plot the profile log-likelihood function manually.

### Logit Model Example

Suppose we have binomial data with a single covariate  $x$  with  $N$  levels, so that the number of successes at  $x_i$  is  $Y_i \sim \text{Binomial}(n_i, \pi_i)$ ,  $i = 1, \dots, N$ . We fit the logit model

$$\log \left( \frac{\pi_i}{1 - \pi_i} \right) = \beta_1 + \beta_2 x_i.$$

The log-likelihood function is then

$$\begin{aligned} \log L(\beta_1, \beta_2) &= \sum_{i=1}^N \left[ y_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + n_i \log(1 - \pi_i) + \log \binom{n_i}{y_i} \right] \\ &= \sum_{i=1}^N \left[ y_i(\beta_1 + \beta_2 x_i) + n_i \log \left( \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i)} \right) + \log \binom{n_i}{y_i} \right]. \end{aligned}$$

The MLE's  $\hat{\beta}_1$  and  $\hat{\beta}_2$  can be computed from `glm`.

Consider calculating a profile likelihood confidence interval for  $\beta_2$ . The profile likelihood function is

$$L_1(\beta_2) = \max_{\beta_1} L(\beta_1, \beta_2).$$

We'll work with the profile log-likelihood function

$$\log L_1(\beta_2) = \max_{\beta_1} \log L(\beta_1, \beta_2).$$

The profile log-likelihood maximization requires us to fix a value of  $\beta_2$  and maximize the log-likelihood function over  $\beta_1$ . We could do this in R using the `optimize` function (this is the R function recommended for maximization for a function of one variable; `optim` is best for functions of two or more variables). However, there is a way to get the `glm` function to do it by using the `offset` option in the `glm` command.

## Offset

To compute the profile log-likelihood function, we first need to be able to maximize over  $\beta_1$  the log-likelihood function for the logit model with a fixed value of  $\beta_2$ , say  $\beta_{20}$ . What that amounts to is fitting a logit model where the linear predictor  $\eta_i$  for bid  $x_i$  is

$$\eta_i = \beta_1 + \beta_{20}x_i.$$

This isn't quite a GLM. There is one parameter to be estimated,  $\beta_1$ , but there is an extra term, not constant across groups, but without a coefficient to be estimated ( $\beta_{20}$  is fixed). The second term is called an offset in a GLM: a term to be added to the linear predictor with known coefficient 1 rather than an estimated coefficient. There is an `offset` option in the `glm` command in R. In this example, we would fit a constant model ( $y/n \sim 1$ ) but with an offset equal to  $\beta_{20}x_i$ . This is illustrated in the example below.

## Goose permit data

Consider the goose permit data from Bishop and Heberlein (*Amer. J. Agr. Econ.* 61, 1979) where 237 hunters were each offered one of 11 cash amounts (bids) ranging from \$1 to \$200 in return for their permits. The logit model was fit with `log(bid)` as the explanatory variable; that is, the probability  $\pi_i$  that a permit holder would accept bid amount  $x_i$  is modeled as

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_1 + \beta_2 \log(x_i).$$

The results of the `glm` fit and the profile likelihood confidence intervals from the built-in R function `confint` are also reported.

```
> bid <- c(1,5,10,20,30,40,50,75,100,150,200)
> n <- c(31,29,27,25,23,21,19,17,15,15,15)
> y <- c(0,3,6,7,9,13,17,12,11,14,13)
> m1 <- glm(y/n~log(bid),weights=n,family=binomial)
> m1
Call:  glm(formula = y/n ~ log(bid), family = binomial, weights = n)
Coefficients:
(Intercept)      log(bid)
      -4.453         1.296
Degrees of Freedom: 10 Total (i.e. Null);  9 Residual
Null Deviance:      119.3
Residual Deviance: 10.63      AIC: 44.43

> confint(m1)
```

```
Waiting for profiling to be done...
```

```
          2.5 %    97.5 %  
(Intercept) -5.783372 -3.317595  
log(bid)      0.978029  1.667461
```

```
> confint(m1,level=.99)
```

```
Waiting for profiling to be done...
```

```
          0.5 %    99.5 %  
(Intercept) -6.2463329 -2.996867  
log(bid)      0.8875305  1.796307
```

The following code computes the profile log-likelihood function for  $\beta_2$ . It also computes the cutoff value  $\log L(\hat{\beta}_1, \hat{\beta}_2) - \chi_{1-\alpha}^2(1)/2$  for a 95% confidence interval and puts it on the plot. The 95% profile likelihood confidence interval is the set of points where the profile log-likelihood function is above the cutoff.

```
> # Compute profile log-likelihood function for beta_2  
> # logLik is a built-in R function to compute log-likelihood of model  
> k <- 200  
> b2 <- seq(.7,2,length=k)  
> w <- rep(0,k)  
> for(i in 1:k){  
+   mm <- glm(y/n~1,offset=b2[i]*log(bid),weights=n,family=binomial)  
+   w[i] <- logLik(mm)  
+ }  
  
> plot(b2,w,type="l",ylab="Profile log-likelihood",cex.lab=1.3,cex.axis=1.3)  
> abline(h=logLik(m1)-qchisq(.95,1)/2,lty=2)
```

If we want to find the 95% profile likelihood confidence limits for  $\beta_2$  without visually approximating them from the graph, we could examine the values of **b2** and **w** (use `cbind(b2,w)`) in R to find the values of **b2** where **w** is approximately equal to the cutoff. We could do a finer search in the neighborhood of each candidate value to get a more accurate approximation. Alternatively, we could have R search for the confidence limits using the function `uniroot` which finds a zero of a function of one variable. It can only find one root at a time so we can have it search for the lower limit and upper limit separately.

In the code below, **f** is the profile log-likelihood minus the cutoff value. The zero's of **f** are the confidence limits for  $\beta_2$ . We get 95% confidence limits of about .9780 and 1.667 which matches the output from `confint` to 3 decimal places.

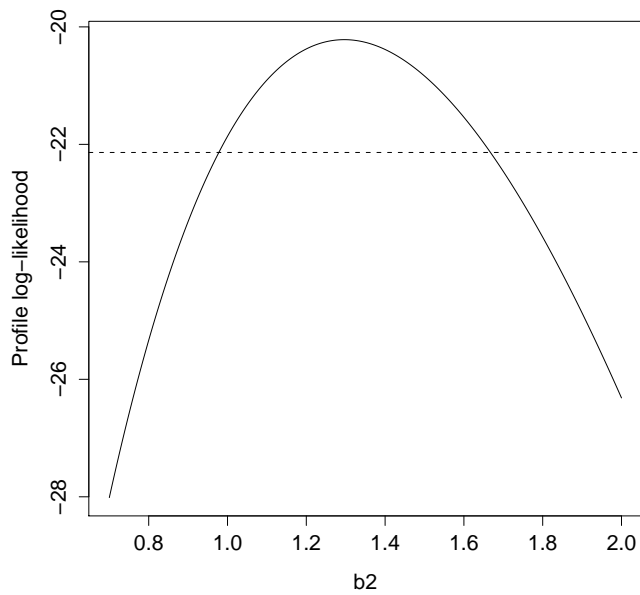


Figure 1: Profile log-likelihood function with 95% confidence limits for  $\beta_2$  in goose data logit model

```

> f <- function(b2,n,x,y,maxloglik){
+   mm <- glm(y/n~1,offset=b2*x,weights=n,family=binomial)
+   logLik(mm) - maxloglik + qchisq(.95,1)/2
+ }
>
> uniroot(f,c(.8,1.1),n=n,x=log(bid),y=y,maxloglik=logLik(m1))
$root
[1] 0.9779875
$f.root
'log Lik.' 1.084503e-05 (df=1)
$iter
[1] 5
$estim.prec
[1] 8.564891e-05
> uniroot(f,c(1.5,1.8),n=n,x=log(bid),y=y,maxloglik=logLik(m1))
$root
[1] 1.667439
$f.root
'log Lik.' -0.0001146866 (df=1)
$iter

```

[1] 4

\$estim.prec

[1] 6.103516e-05

## Profile likelihood confidence intervals for functions of parameters

Profile likelihood confidence intervals have the nice property that they are invariant under monotonic transformation of a parameter; that is, if  $(c_1, c_2)$  is a profile likelihood confidence interval for a parameter  $\theta$ , then  $(g(c_1), g(c_2))$  is a profile likelihood confidence interval for  $g(\theta)$ . For example, in the notes for Chapter 7, I showed how to obtain a confidence interval for the mean of the log-tolerance distribution,  $-\beta_1/\beta_2$  using the delta method and then for the median of the tolerance distribution,  $\exp(-\beta_1/\beta_2)$ . There were two alternatives for the latter: derive the SE from the delta method and form a normal-based confidence interval, or form a normal-based confidence interval for  $-\beta_1/\beta_2$  and transform the endpoints to a confidence interval for  $\exp(-\beta_1/\beta_2)$ . With profile likelihood confidence intervals, there isn't this issue, although calculating the CI's may be a little more work, which I'll now illustrate.

### Goose example, continued

To calculate a profile likelihood CI for  $\alpha = -\beta_1/\beta_2$ , we reparameterize the logit model in terms of the location parameter  $\alpha$  and the scale parameter  $\kappa = 1/\beta_2$ . The link function is

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -\frac{\alpha}{\kappa} + \frac{1}{\kappa}x_i.$$

The log-likelihood function is then

$$\begin{aligned}\log L(\alpha, \kappa) &= \sum_{i=1}^N \left[ y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) + \log\left(\frac{n_i}{y_i}\right) \right] \\ &= \sum_{i=1}^N \left[ y_i \left(-\frac{\alpha}{\kappa} + \frac{1}{\kappa}x_i\right) + n_i \log\left(\frac{1}{1 + \exp(-\alpha/\kappa + (1/\kappa)x_i)}\right) + \log\left(\frac{n_i}{y_i}\right) \right].\end{aligned}$$

The profile likelihood function for  $\alpha$  is

$$L_1(\alpha) = \max_{\kappa} L(\alpha, \kappa).$$

In order to calculate  $L_1(\alpha_0)$  for some fixed value  $\alpha_0$ , we are fitting a binomial GLM with link function

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -\frac{\alpha_0}{\kappa} + \frac{1}{\kappa}x_i = \beta(x_i - \alpha_0)$$

where  $\beta = 1/\kappa$  is the only parameter to be estimated. This is a logit model with no intercept and a single covariate,  $x_i - \alpha_0$  and we can use the `glm` command to fit it (note: put a `-1` in the formula for the `glm` to omit the intercept). The R script to compute the profile log-likelihood function and a confidence interval for  $\alpha$  and for  $\exp(\alpha)$  is:

```

> # Profile log-likelihood function for alpha = -beta_1/beta_2
> k <- 200
> alpha <- seq(3,4,length=k)
> w <- rep(0,k)
> for(i in 1:k){
+   xx <- log(bid)-alpha[i]
+   mm <- glm(y/n~xx-1,weights=n,family=binomial)
+   w[i] <- logLik(mm)
+ }
plot(alpha,w,type="l",ylab="Profile log-likelihood",cex.lab=1.3,cex.axis=1.3)
abline(h=logLik(m1)-qchisq(.95,1)/2,lty=2)

# Find confidence limits for alpha
> f2 <- function(mu,n,x,y,maxloglik){
+   xx <- x-mu
+   mm <- glm(y/n~xx-1,weights=n,family=binomial)
+   logLik(mm) - maxloglik + qchisq(.95,1)/2
+ }
> u1 <- uniroot(f2,c(3.1,3.4),n=n,x=log(bid),y=y,maxloglik=logLik(m1))
> u2 <- uniroot(f2,c(3.6,3.9),n=n,x=log(bid),y=y,maxloglik=logLik(m1))
> c(u1$root,u2$root) # 95% confidence interval for mu
[1] 3.175975 3.691953
> exp(c(u1$root,u2$root)) # 95% confidence interval for median value of goose permit
[1] 23.95016 40.12311

```

The 95% profile likelihood confidence interval for the median value of a permit (\$23.95 to \$40.12) is very close to that obtained by transforming the delta method confidence interval for  $\alpha$  (\$24.08 to \$39.96).

It will not always be the case that the profile likelihood function for a parameter can be obtained using the `glm` command with an offset or some other “trick.” Sometimes, it will be necessary to do the maximization directly using `optimize` or `optim`.

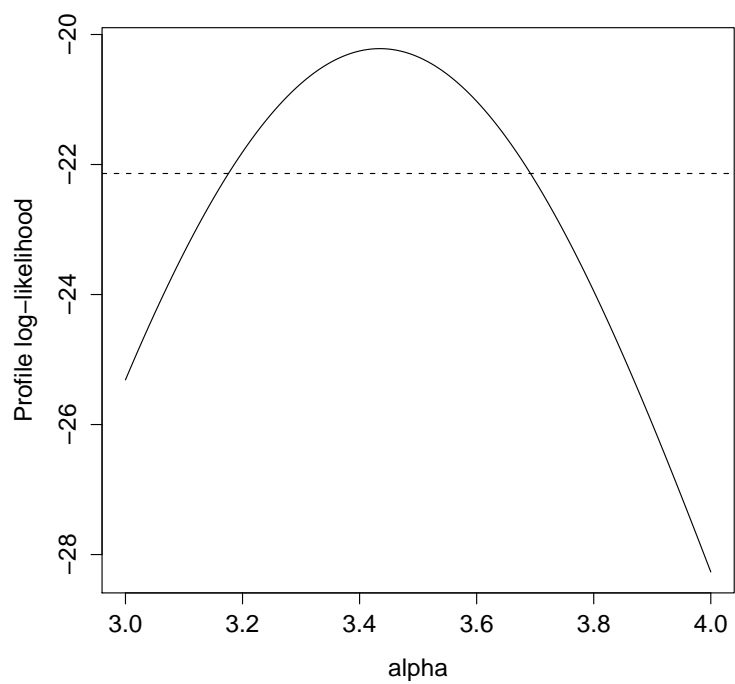


Figure 2: Profile log-likelihood function with 95% confidence limits for  $\alpha$  in goose data logit model